

Uma Análise Quantitativa dos Repositórios Digitais Brasileiros

Ana Lameira, Franciane Oliveira, Thaisa Ramos & Washington Carvalho-Segundo

30 de Outubro de 2016

Objetivos

Analisar a distribuição do quantitativo de documentos disponíveis nos repositórios digitais brasileiros, nos quesitos de:

1. Concentração de documentos por região, inserindo-se Portugal na comparação da produção científica em acesso aberto disponível em RIs;
2. Tipo de repositório: institucional, biblioteca digital genérica, base de teses e dissertações eletrônicas, temático, objetos educacionais, preservação de acervo histórico, trabalhos de conclusão de cursos acadêmicos;
3. Tipo de mídia suporte: texto, imagem, vídeo, código fonte *etc*;
4. Versões da plataforma DSpace: 1.x, 3.x, 4.x, 5.x e 6.x.

Metodologia

- Efetuaram-se pesquisas restritas ao Brasil no portal de buscas Google (2016), pelos termos: “DSpace”, “TEDE”, “Nou-Rau”, “Fedora”, “Bepress”.
- Estes são os principais softwares utilizados no mundo e no Brasil para construção de repositórios digitais de acordo com a lista de fontes de coleta do portal oasisbr (2016) e com o ROAR (2016).
- As pesquisas foram exportadas no formato CSV por meio do addon SeoQuake do navegador Mozilla Firefox.
- Uma seleção manual foi realizada para captar apenas URLs relativas a repositórios digitais.
- A lista obtida foi somada ao conjunto das fontes de coleta de RIs e repositórios de teses e dissertações do oasisbr.
- Para cada repositório da lista, realizou-se o levantamento do software utilizado, versão, número de registros armazenados, localização geográfica, e o tipo de repositório (institucional, de teses e dissertações *etc*).
- Além disso, foram desenvolvidos scripts na linguagem de programação Python que identificaram, nos repositórios institucionais e nas bases de teses e dissertações, qual o tipo de mídia armazenada (texto, áudio, vídeo *etc*).
- Para que fosse possível obter uma taxa de densidade geográfica dos documentos disponibilizados nos repositórios, foram recolhidos também dados de projeções demográficas das regiões brasileiras nas tabelas do IBGE (2016), além dos dados da população de doutores envolvidos com atividades de pesquisa descritos na Plataforma Lattes (2016), e do estudo realizado por Neto et al. (2014) para obtenção de uma estimativa do número de pesquisadores doutores residentes em Portugal.

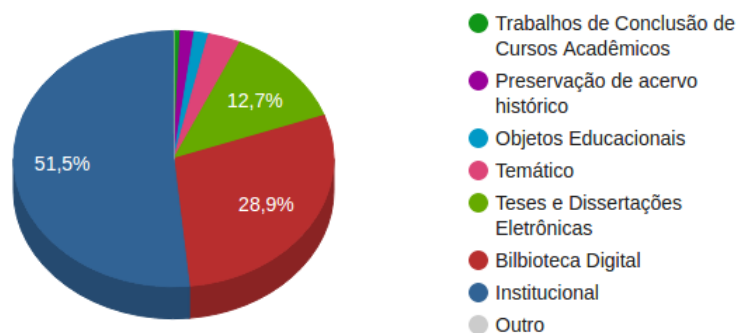


Figura 1: Documentos por tipo de repositório

Resultados

Foram identificados 186 repositórios distribuídos por 147 instituições, que somaram pouco mais de 1 milhão e 900 mil documentos. As análises efetuadas foram relacionadas sempre ao número de documentos agrupados em cada classe, que são as seguintes:

1. tipo de repositório,
2. concentração geográfica,
3. plataforma e versão do software utilizado, e
4. formato da mídia armazenada.

No quesito "tipo de repositório", as bases foram agrupadas em: biblioteca digital, catálogo, repositório comum, diretório, repositório institucional, objetos educacionais, preservação de acervo histórico, repositório temático, base de teses e dissertações e repositório de trabalhos de conclusão de curso. Constatou-se que 51,5% dos documentos estavam abrigados em repositórios institucionais, 28,9% em bibliotecas digitais, 12,7% em bases exclusivas de teses e dissertações, e o percentual restante (6,9%) está distribuído pelos demais tipos de repositórios (Figura 1).

Com relação a concentração geográfica dos documentos nas regiões brasileiras, e levando-se em conta o índice da razão entre número de documentos, das bases em geral, por cada 1000 habitantes, foi obtido o gráfico da (Figura 2), que mostra que há uma grande disparidade entre as regiões norte/nordeste e centro-oeste/sudeste/sul. Destaca-se a produção disponível em repositórios digitais na região centro-oeste.

Para comparar as regiões a nível de produção disponível em repositórios institucionais, considerou-se somente a população de doutores envolvidos com atividades de pesquisa e (fonte: Plataforma Lattes (2016)) e o número de doutores residentes em Portugal (Fonte: NETO et al. 2014). Mediu-se o total de documentos portugueses indexados pelo portal RCAAP (2016) e a razão entre o número de documentos abrigados em repositórios institucionais e o número de doutores de cada região ou país. O resultado pode ser observado na Figura 3, com destaque para as regiões brasileiras centro-oeste, sul, sudeste e Portugal.

Os formatos encontrados foram classificados em: imagem, texto, planilha, vídeo, código fonte, áudio, tipo indeterminado e apresentação. Para RIs, observou-se a seguinte classificação: 75,9% são textos, 12,6% são imagens, 10,7% indeterminados e os 0,8% restantes são distribuídos pelos demais tipos. Já entre as bases de teses e dissertações, 90,3% foram classificados como tipo texto, 7,3% como código (html, rdf etc) e os restantes 2,4% estão distribuídos pelos demais tipos.

Com relação as plataformas utilizadas, 82,4% dos documentos são abrigados em bases que utilizam DSpace, 5,74% estão em bases de desenvolvimento próprio, 3,63% estão em sistemas TEDE, 3,59% em Nou-Rau, 3,14% em Joomla, 1,37% em Aleph e 0,08% em SophiA.

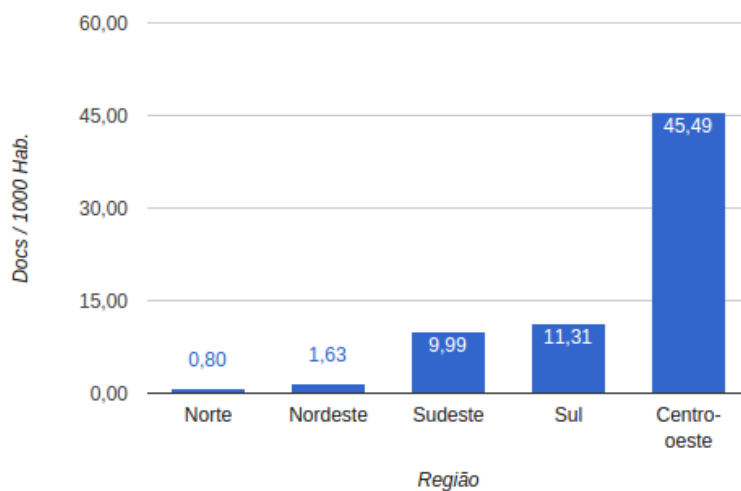


Figura 2: Documentos disponíveis e repositórios digitais por região, para cada 1000 habitantes

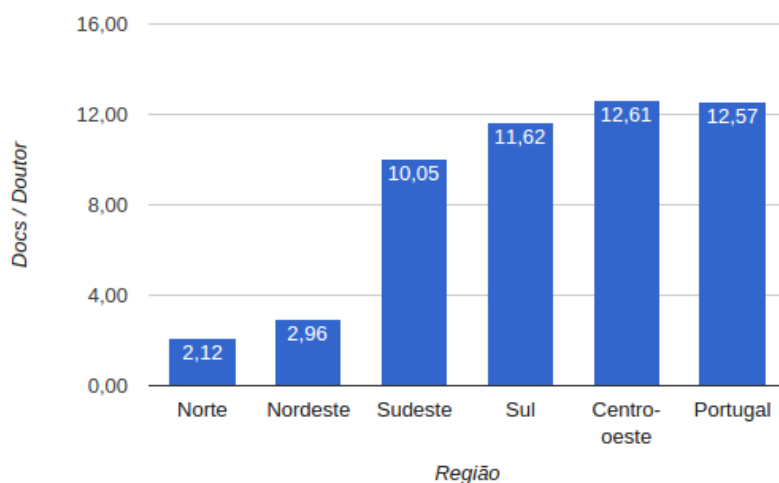


Figura 3: Documentos disponíveis em repositórios institucionais por região, na razão da população de pesquisadores doutores

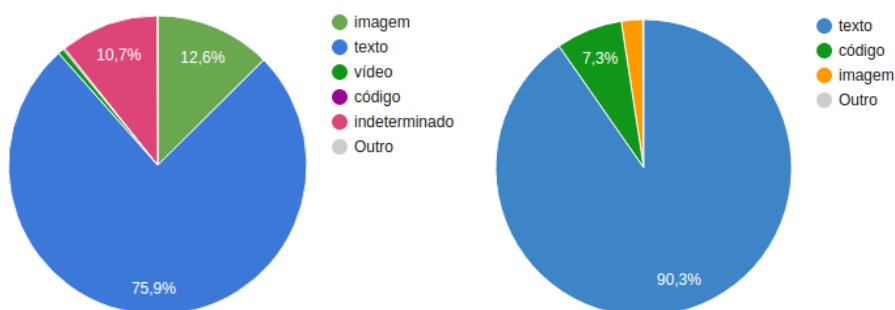


Figura 4: Documentos por mídia de suporte em repositórios institucionais e bases de teses e dissertações

Conclusões e trabalhos futuros

- Repositórios institucionais abrigam metade da produção disponível em RDs;
- Observa-se um surpreendente destaque no quantitativo de documentos disponibilizados na região centro-oeste, tanto com relação a RDs quanto a RIs;
- Em contrapartida, as regiões norte e a nordeste não conseguiram ainda imprimir um volume condizente com seus números de habitantes e populações de pesquisadores doutores;
- DSpace é, de longe, a plataforma mais utilizada para construção de repositórios digitais;
- Conforme esperado, a maioria dos documentos armazenados em RIs e bases de teses e dissertações está ainda em mídias do tipo texto, salvo a exceção de que em particular para TDs observou-se expressivos 7,3% no formato de código fonte.
- Pretendemos estender a investigação relativa ao tipo da mídia armazenada para que no futuro tenhamos também uma análise quantitativa sobre repositórios de dados de pesquisa.

Referências

- [1] HEERY, Rachel; ANDERSON, Sheila. Digital repositories review. 2005. Disponível em: <<http://opus.bath.ac.uk/23566/2/digital-repositories-review-2005.pdf>>. Acesso em: abril de 2016.
- [2] LEITE, Fernando César Lima et al. Como gerenciar e ampliar a visibilidade da informação científica brasileira: repositórios institucionais de acesso aberto. Fernando César Lima Leite, 2009.
- [3] NETO, Catarina; MEIRELES, Gonçalo; BARROCA, Ana. A empregabilidade e perspectivas de carreira dos doutorados nas organizações portuguesas. *Investigação e Intervenção em Recursos Humanos*, v. 1, n. 1, 2014.